

OPEN AND SHUT?

Sunday, May 04, 2014

Interview with Kathleen Shearer, Executive Director of the Confederation of Open Access Repositories

In October 1999 a group of people met in New Mexico to discuss ways in which the growing number of "eprint archives" could co-operate.



Kathleen Shearer

Dubbed the **Santa Fe Convention**, the meeting was a response to a new trend: researchers had begun to create subject-based electronic archives so that they could share their research papers with one another over the Internet. Early examples were **arXiv**, **CogPrints** and **RePEc**.

The thinking behind the meeting was that if these distributed archives were made interoperable they would not only be more useful to the communities that created them, but they could "contribute to the creation of a more effective scholarly communication mechanism."

With this end in mind it was decided to launch the Open Archives Initiative (OAI) and to develop a new machine-based protocol for sharing metadata. This would enable third party providers to harvest the metadata in scholarly archives and build new services on top of them. Critically, by aggregating the metadata these services would be able to provide a single search interface to enable scholars interrogate the complete universe of eprint archives as if a single archive. Thus was born the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH). An early example of a metadata harvester was **OAIster**.

Explaining the logic of what they were doing in *D-Lib Magazine* in 2000, Santa Fe meeting organisers **Herbert Van de Sompel** and **Carl Lagoze** wrote, "The reason for launching the Open Archives initiative is the belief that interoperability among archives is key to increasing their impact and establishing them as viable alternatives to the existing scholarly communication model."

As an example of the kind of alternative model they had in mind Van de Sompel and Lagoze cited a recent proposal that had been made by **three Caltech researchers**.

Today eprint archives are more commonly known as open access repositories, and while OAI-PMH remains the standard for exposing repository metadata, the nature, scope and function of scholarly archives has broadened somewhat. As well as subject repositories like **arXiv** and **PubMed Central**, for instance, there are now thousands of institutional repositories. Importantly, these repositories have become the primary mechanism for providing **green open access** — i.e. making publicly-funded research papers freely available on the Internet. Currently **OpenDOAR** lists over 3,600 OA repositories.

Work in progress

Fifteen years later, however, the task embarked upon at Santa Fe still remains a work in progress. Not only has it proved hugely difficult to persuade many researchers to make use of repositories, but the full potential of networking them has yet to be realised, not least because many repositories do not attach complete and consistent metadata to the items posted in them, or they only provide the metadata for a document, not the document itself. As a consequence, locating and accessing content in OA repositories remains a hit and

- Home
- About Richard Poynder
- Blog: *Open and Shut?*
- The State of Open Access
- The Basement Interviews
- Open Access Interviews
- Essays on Open Access
- Archive

Blogs, News & More:

Interview 1: Richard Poynder

Interview 2: Richard Poynder

Interview 3: Richard Poynder

DOAJ

BASE

LENS

Digital Koans

LSE Impact Blog

Heather Morrison

The Scholarly Kitchen

Open Access India

PLoS Blogs

Redalyc

SPARC Europe

IP Watchdog



Search This Blog

Popular Posts



Open Access: "Information wants to be free?" (A print version of this eBook is

available here) Earlier this year I was invited to discuss with Georgia Institute of Technology librarians...



PLOS CEO Alison Mudditt discusses new OA agreement with the University of

California

Tweets by @RickyPo



Richard Poynder
@RickyPo

Ministerial interference is an attack on academic freedom and Australia's literary culture
theconversation.com/ministerial-in-... via [@ConversationEDU](https://twitter.com/ConversationEDU)

1h



Richard Poynder
@RickyPo

Confessions of a 'human guinea pig': Why I'm resigning from Moderna vaccine trials
statnews.com/2022/01/04/

The State of Open Access
Predatory publishing
Institutional Repositories
Green OA Gold
OA Self-archiving
Copyright
Basement Interviews
OA Interviews

OA Essays Open
Access in Serbia
Open Access in
India Open Access in
 Egypt **ScienceDirect**
Open Access
in California
OA in Latin
America Open
Access in the Humanities
MDPI Preprints Selecting
 Reviewers **Global**
Research
Council OA Big
Deal Open
Notebook Science
Elsevier Gates
Foundation OA in
South Africa OA in
France SSRN OA &
the Humanities
 Timothy Gowers Harold
 Varmus **Peter**
Suber OA in Poland
OA Embargoes
Big Deal Finch
Report Jeffrey Beall
ALPSP OA
Mandates PLOS
Peer Review
Springer BioMed
Central Free
Software Digital
Preservation Dove
Medical OA in Russia
Radical OA Almost OA
HEFCE Frontiers

miss affair, and while many researchers now turn to Google and Google Scholar when looking for research papers, Google Scholar has not been as receptive to indexing repository collections as OA advocates had hoped.

For scholars, the difficulties associated with accessing papers in repositories is a continuing source of **frustration**. Meanwhile, critics of green OA argue that the severe shortage of content in them means that any hope of building an effective network of OA repositories is a lost cause anyway.

For their part, conscious that green OA poses a potential threat to their profits, publishers have responded to the growing calls for open access by offering pay-to-publish gold OA journals as an alternative.

It was against this background that in 2012 the Finch Committee **concluded** that in order for the UK to make an effective transition to OA “a clear policy direction should be set towards support for publication in open access or hybrid journals, funded by APCs, as the main vehicle for the publication of research.”

Explaining the decision to prioritise gold OA, Finch argued that repositories had failed to deliver on their promise. “Despite the best efforts of repository managers and librarians ... rates of deposit and usage of published materials remain fairly low; and a number of issues will need to be addressed if institutional repositories are to fulfil a bigger and more effective role in the research communications landscape.”

For that reason, Finch added, repositories should in future be viewed as being merely “complementary to formal publishing, particularly in providing access to research data and to grey literature, and in digital preservation”

The Finch Report proved highly controversial, particularly when Research Councils UK (RCUK) responded by introducing a new gold-preferred OA Policy conforming to its recommendations. Many OA advocates in particular felt betrayed.

But we need to ask: did Finch have a point?

We should not doubt that huge challenges remain in getting content into repositories. However, the whys and wherefores of this have been well rehearsed elsewhere, so we won't dwell on them here.

Instead, let's consider the current state of the repository infrastructure, particularly with regard to interoperability and discoverability. Why, for instance, do many repositories not expose adequate metadata? Why do they sometimes provide just the metadata and not the full text? When will the sophisticated search functionality that researchers need become standard in repositories? Will it? And what new developments might help here? More generally, what does the future hold for the OA repository?

Investing for the long term

Who better to put these questions to than Kathleen Shearer, Executive Director of the Confederation of Open Access Repositories (COAR)? Launched in **October 2009**, COAR's mission is to “enhance the visibility and application of research outputs through a global network of open access digital repositories” and its membership currently includes over 100 institutions from around the world.

Reading Shearer's replies below one has to conclude that there is much still to be done. Scholars and scientists will therefore clearly need to be patient. And while new repositories are constantly being created, and existing ones improved (as are cross-repository search services like **BASE**), the truth is that if the vision articulated in New Mexico fifteen years ago is to be fully realised the research community is going to have to invest a great deal more time, effort and money to developing its repositories.

But should it? Now that most if not all scholarly publishers offer gold OA is further investment in repositories justified?

Shearer believes it is – for two reasons. First, she says, wide-scale take up of green OA would contain publishers' prices; second, the time has in any case come for the research community to take back control of the scholarly communication system, and repositories will be vital in doing that.

The Public Library of Science (PLOS) and the University of California (UC) have today announced a two-year agreement designed to make...



P2P: The very core of the world to come
In the first part of this interview Michel

Bauwens , the creator of The Foundation for P2P Alternatives , explained why he believes the var...



The OA Interviews: Taylor & Francis' Deborah Kahn discusses Dove

Medical Press

Please note the postscript to this interview here The open-access publisher Dove Medical Press has a controversial past and I have writ...



The Open Access Interviews: Publisher MDPI Headquartered in Basel,

Switzerland, the Multidisciplinary Digital Publishing Institute, or more usually MDPI , is an open access publisher...



Community Action Publishing: Broadening the Pool

We are today seeing growing dissatisfaction with the pay-to-publish model for open access. As this requires authors (or their funders or ins...



Copyright: the immovable barrier that open access advocates underestimated

In calling for research papers to be made freely available open access advocates promised that doing so would lead to a simpler, less cos...



The Open Access Interviews: OMICS Publishing Group's Srinu

Babu Gedela

***Update: On August 26th 2016, the US government (Federal Trade Commission) announced that it has charged OMICS with making false claims, ...



Robin Osborne on the state of Open Access: Where are we, what still needs to be done?

As Shearer puts it, “[T]he Green Road is key. We must collectively build and maintain a global system of repositories. It introduces competition into the system and will act as an important deterrent to arbitrary price increases by publishers.”

She adds, “It will also demonstrate the important role that institutions play in the stewardship of research outputs. To that end, institutions should devote more resources to their repository operations in order to improve repository services and increase the size of their collections.”

As I read it, the promise is that any investment made in OA repositories today will more than pay for itself in the long term.

The interview begins

RP: Can you say who you are, where you are based and what role you play within COAR?

KS: I am the Executive Director of COAR and I am based in Montreal, Canada, although the COAR office is located in Göttingen, Germany. I have been working in the area of open access and digital repositories for about a dozen years now, mainly in the Canadian context as a consultant and a research associate with the [Canadian Association of Research Libraries](#). In June 2013, I became the Executive Director of COAR.

RP: Briefly, what is COAR, how is it funded, and what is its purpose?

KS: COAR, the Confederation of Open Access Repositories, is an association of repository initiatives with an international membership.

We have over 100 members in 35 countries around the world. Our members come from a variety of communities including universities/libraries, research institutions, funding agencies, intergovernmental organizations and government departments – any organization that may have an interest in repository development and wants to be connected with the international community.

COAR’s mission is to raise the visibility of research outputs through a global network of repositories. We are active on two levels: (1) At the practical level, we support communities of practice around areas of importance for our members mainly in terms of best practices, interoperability and monitoring trends in the repository landscape and (2) At the strategic level, we aim to facilitate greater alignment of regional and national repository networks around the globe.

COAR is funded mainly through membership fees, although we receive in-kind support for our office space from the [University of Göttingen](#) and some partnership funding as well.

We are quite a light-weight organization with about 1.5 full time positions in total and an Executive Board chaired by [Norbert Lossau](#), Vice-President of the University of Göttingen. Most of our activities are undertaken by the active participation of our members.

RP: The mission of COAR, you said, is to “raise the visibility of research outputs through a global network of repositories”. I think it might help if we tried to clarify what this means in practice. In other words, what do we mean by repository here, and what role exactly do we expect that repository to play? Are we talking about a global network of institutional repositories, or does repository here encompass more than that (i.e. central subject-based repositories like [PubMed Central](#) and [arXiv](#) too, and perhaps other content management systems and databases?)

Likewise, should we assume the role of the repository remains as it was originally conceived – a tool to support green OA by providing a place where papers published in subscription journals can be self-archived in order to ensure that free copies are always available outside the subscription paywall?

Or do we assume that the repository can now also act as a publishing platform on which institutions can publish their own journals – as currently planned, for instance, by University College London?

One of a series exploring the current state of Open Access (OA), the Q&A below is with Robin Osborne , Professor of Ancient History a...



The OA Interviews:
Frances Pinter
In 2012 serial entrepreneur Frances Pinter

founded a new company called Knowledge Unlatched (KU). The goal, she explained in 2013, was ...

Blog Archive

2020 (4)
2019 (7)
2018 (20)
2017 (18)
2016 (14)
2015 (18)
2014 (13)
December (2)
September (1)
August (1)
June (4)
May (2)
Working for a phase transition to an open commons-...
Interview with Kathleen Shearer, Executive Directo...

April (1)
March (1)
February (1)

2013 (32)
2012 (43)
2011 (22)
2010 (20)
2009 (22)
2008 (14)
2007 (9)
2006 (27)
2005 (31)
2004 (2)

Followers

Alternatively, perhaps the assumption is that today the repository should be viewed as little more than what the [Finch Report](#) assumed it to be: something “complementary to formal publishing, particularly in providing access to research data and to grey literature, and in digital preservation” (A model that assumes open access is provided by means of gold rather than green OA)?

KS: Repositories are evolving and play a number of roles. At their core, a ‘repository’ could be theoretically defined as a set of services that provide open access to research outputs (along the lines of Cliff Lynch’s [original definition in 2003](#)). However, in practice, repository services and infrastructures are diverse and there is a lot of overlap with other systems. Perhaps most significantly, practices and technologies are changing quickly, making it a challenge to concretely define their services. My feeling is that we need to be flexible in the way we conceptualize repositories.

In terms of COAR, we are a community brought together by a set of shared principles and common practices rather than by a narrowly delineated concept of repository. So yes, we would include disciplinary repositories and content management systems (if they provide open access to full text) in our global network.

In terms of a complement to formal publishing, I expect that traditional publishing will soon be going through some pretty big transitions, likely some very disruptive changes. I agree with [Dominique Babini](#), Jean-Claude Guédon and others that we should aim for a basic, open, and interoperable system that is free to both access and contribute to. Value-added services by publishers and others can be built on top of this content.

One way of thinking about repositories is that they represent an institutional commitment to the stewardship of research outputs. In this sense, they address two important problems in the current system: sustainability and stewardship.

I believe institutions should assume greater responsibility for managing, providing access and preserving the content created through research. It will alleviate some of the inflationary aspects of scholarly publishing and enable us to have more influence on future directions. This was the traditional mission of libraries in the print world, which has been somewhat lost in the transition to digital content. How this plays out in terms of models will likely vary according to content type, discipline, and region.

Interoperability

RP: *I would like to focus on the issue of interoperability. I am aware of a number of current initiatives devoted to getting institutional repositories to interact/interoperate, including DRIVER, DRIVER II, euroCRIS, OpenAIRE and no doubt there are others too. How do these various initiatives fit together (do they?), and why are there so many initiatives that – to the layperson at least – might seem to be duplicating effort?*

KS: There are several initiatives that have evolved from different requirements, regions, and with differing aims.

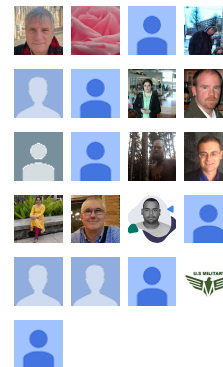
DRIVER and DRIVER II were European Commission-funded projects to support the implementation of repositories in EU countries. The aim was to have repositories adopt common guidelines for organizing their content so they could be harvested and searched through the DRIVER search service.

OpenAIRE has built upon work of DRIVER to implement further standards that enable the European Commission to track the open access research output they fund. Each of these three projects required some level of interoperability between participating repositories.

There are similar initiatives in other regions, such as [La Referencia](#) in Latin America and SHARE in the US that will also require some level of interoperability across those repository networks.

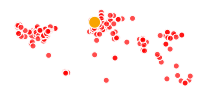
COAR is a forum whereby all of these regional initiatives can work together to identify issues in common and, where appropriate, agree on standardized practices. COAR will be intensifying efforts in this area and has just launched an [initiative](#) to address some of the differences between repository networks that are evolving.

Followers (117) [Next](#)



[Follow](#)

802 Pageviews
Dec. 04th - Jan. 04th



Open & Shut? by [Richard Poynder](#) is licensed under a [Creative Commons Attribution-NonCommercial-NoDerivs 2.0 UK: England & Wales License](#).

Permissions beyond the scope of this license are available [here](#)

EuroCRIS is a European association that is looking at interoperability between research administrative systems. The objective of these systems is to manage and report on research activities. Unlike repositories, CRIS systems do not usually manage full text content.

We have seen in the last few years some merging between CRIS systems and repositories, with some repositories being integrated with CRIS's, or at least interoperability between repositories and CRIS.

COAR has also been working with EuroCRIS to identify strategies for greater interoperability between research administration systems and repositories.

RP: The concept of networking repositories dates back at least to 1999, and the [Santa Fe Convention](#). I believe it was in the wake of the Santa Fe meeting that the OAI-PMH protocol was developed. However, I assume that both the thinking and the technology have developed somewhat since then.

As I understand it, for instance, OAI-PMH was based on the principle that services would be developed to harvest metadata from repositories in order to aggregate their holdings and provide a centralised discovery service. I guess this assumed that records in repositories would consist of metadata but not the full text (so the goal presumably was to signal where papers were held, not to provide direct access to them).

I would think that the emphasis today is more on providing direct access to full-text documents not just their metadata. Briefly, therefore, can you say how thinking has developed since 1999, and how the technologies and protocols have changed to reflect this?

KS: OAI-PMH was developed on the principle that a service would harvest the metadata record that would then point the user back to the full text content in the repository. So in that sense it does facilitate access to the full text, but without having to aggregate the content into a central archive.

OAI-PMH is still the common denominator for metadata exposure in repositories and it remains standard practice for cross-repository search services to harvest metadata and then point the user back to the repository to access the full text. Full text harvesting is much more demanding, requiring large storage space to house the content in a central location and there are other technical challenges attached to full text harvesting.

The disadvantage of metadata harvesting is that the search services are based on the metadata supplied by the repositories, which isn't always comprehensive, complete or consistent. COAR aims to improve the current situation by identifying and encouraging the adoption of common standards and metadata globally. However, for better discoverability, and especially for other services such as text mining, using full text search is highly desirable.

In terms of discovery, repository managers have found that most users find the content in repositories through search engines such as Google and Google Scholar, not from metadata harvesting services or by directly searching the repository. Therefore, the repository community has put significant efforts into exposing their content to commercial search engines through various optimization techniques.

Beyond discoverability, there are other areas of repository networking and interoperability, like content transfer, usage data, etc. where new technologies and standards/protocols have been created. COAR is a forum whereby interoperable practices can be agreed upon globally.

Full text

RP: You say that it remains standard practice for cross-repository search services to harvest metadata and then point back to the full text in the repository, and you said that COAR assumes OA repositories will "provide open access to full text". This would seem to imply that an OA repository always now includes the full-text as well as the metadata (and indeed most people would presumably expect that of an OA repository).

However, not all records in OA repositories do provide access to the full-text, and many seem to offer little more than the bibliographic details. Even

a poster child of the OA movement – Harvard’s DASH repository – has been criticised for not providing the full text (e.g. [here](#)). These criticisms were made a few years ago, but DASH does still today contain records without any full-text attached. Moreover, some do not even provide a link to the full-text (and DASH does not seem to have a [RequestCopy Button](#)). When I looked in DASH the other day, for instance, I found (at random) five examples of this (one, two, three, four, five).

I think this cannot be a consequence of publisher embargoes since the articles concerned date back as far as 1993, with the two most recent published five years ago (and in any case the [Harvard OA Policies](#) claim to moot publisher embargoes). Moreover, where in a couple of cases the DASH records do [point](#) to the full-text this is a link to the publisher’s version, where the user is asked to pay for access (\$35 in one case). This cannot be described as OA.

You may not want to comment specifically on DASH, but do you think it problematic when records in OA repositories do not always provide access to the full-text, and maybe don’t even link to a free copy of it? If so, what can/is COAR do/doing to address the situation, in concrete terms?

KS: Ideally, all records in the repository will have the full text attached. However, as you point out, this isn’t always the case. I’m not sure about the specific case of DASH, but this really speaks to the collection policy of the individual repository.

As I said earlier, more and more repositories are now being used to track research output. In that case the objective may be to collect information about all of the publications at the institution, regardless of whether they are open access or not. Still other repositories may be inputting metadata records without the full text as a strategy to encourage authors to upload their documents.

If we look at the OpenAIRE portal as an example, they are currently harvesting 8.4 million records from over 400 sources (mostly repositories, but also open access journal articles). Over 8.2 million of those records are open access. So, I believe that the vast majority of content in repositories is open access, with a small percentage of metadata-only records. The portion of open access, of course, will vary depending on the repository.

In my opinion, the most effective way to improve the proportion of full text in repositories is to continue to advocate for open access policies at funding agencies and institutions around the world. These are the levers that will have a real influence on the policies and practices of the individual repositories. More staffing and resources directed towards repository operations would also help.

RP: You said that rather than searching directly in repositories, or exploiting metadata harvesting services (like [OAster](#) perhaps?), researchers tend to rely on search services like Google and Google Scholar for the discovery of scholarly content in repositories.

Does this mean that the repository community tends today to assume that the research community should rely on mainstream search services, rather than trying to build sophisticated repository search services itself?

If so, I am conscious that OA advocates frequently complain that Google is not supportive enough of their needs, and not as keen to index repository collections as they would like. Would you agree? What is the current situation with regard to mainstream search services like Google, Bing and Yahoo in terms of indexing repositories, and what future developments do you envisage that might improve the situation so far as searching repositories is concerned?

KS: It’s not really about what the repository community believes is the best solution, but rather a practical response to user behaviour.

It would be erroneous to assume all information seekers are the same. However, we do know that even for well-developed disciplinary services, such as PubMed Central and Medline, the majority of users access articles directly from commercial search engines like Google and Google Scholar.

According to my COAR colleague Eloy Rodrigues, Director of the University of Minho Documentation Services, most well developed institutional repositories have about 3/4 of their traffic coming from Google and other generic search

engines. Repository managers take that as very positive sign of the visibility and accessibility of the content in the repository.

In terms of mainstream search engines and Google Scholar there has been ongoing discussion about their efficacy in retrieving scholarly content. It really depends on if you are looking for something you know exists (i.e. you search the title or author's name) or you are searching using key words.

As reported in an article published in the *Online Journal of Public Health Information* (Giustini and Boulos, 2013), "Google Scholar's constantly-changing content, algorithms and database structure make it a poor choice for systematic reviews."

If you are looking for a specific document in a repository and you know the title, the search engine will likely point to it. However, searching by key words, content in repositories are not always high in the rankings.

The problem of visibility is likely even more acute for repositories with non-English content as there does seem to be a bias towards English language content in these search engines.

This will remain an ongoing challenge for repositories as technology continues to change rapidly.

Inherent tension

RP: Certainly there seems to be some disappointment amongst researchers that 15 years after the Santa Fe meeting they still find it extremely difficult, if not impossible, to search effectively in and across OA repositories. I saw this view expressed most recently by Cambridge University chemist Peter Murray-Rust who tweeted, "If libraries provide modern search I'd change my mind; but articles in repos are difficult to discover". His conversation can be viewed [here](#).

Does Murray-Rust have a point? What can you say to convince him that his needs will be met soon? Can you? If so, how will they be met?

KS: There is an inherent tension that exists in the repository community. On the one hand, we aim to make the deposit process as easy as possible so that creators will contribute (or repository staff costs are manageable); on the other hand, we want to assign good quality metadata (which takes time and effort) because we know it will enable greater interoperability and improve discoverability of content. So far, the former has been a greater priority.

There is some truth to Peter Murray-Rust's comments in that complex search services, such as those developed for some discipline-based repositories, require quite a high level of curation, especially for non-textual material. Datasets, for example, need to be accompanied by fairly comprehensive metadata describing them and those metadata elements need to be standardized across each item.

It is a far greater challenge to develop complex searching across numerous repositories containing different disciplines, languages and formats. To facilitate advanced searching in this context, there needs to be interoperability across repositories. COAR has been working on this and this is one of our top priorities; but it takes time to realize this across a very diverse repository landscape.

That being said, there are already a number of cross-repository search services, for example [BASE](#), [CORE](#), and [OpenAIRE](#), which are working to improve the retrieval of content in repositories. They have advanced search options that allow you, for example, to limit your search to publication type, geographic location, publication year and so on. You can't do all of these things in Google Scholar.

OpenAIRE enables users to identify publications related to the projects for which they are funded. These services (and others) will continue to develop and will incorporate more sophisticated tools to improve discovery in the future.

Personally, I can envision a time not too far in the future when more complex search services are built on top of repository networks. What individual repositories should focus on, in my opinion, is ensuring that their content is open, can be indexed, and is attached with the necessary metadata in order to facilitate the development of these services.

RP: From what you have said would it be accurate for me to conclude the following: Users tend to prefer using commercial search engines and Google Scholar for discovering research papers in repositories. However, this is not always the best approach.

We don't yet know exactly what the role of the OA repository will be, nor what form it might eventually take (indeed, repositories will likely take a number of different forms, and play a variety of different roles).

For these reasons it is important that repository managers ensure their content is open, that it has appropriate metadata attached, and that it can be indexed. Doing this will provide sufficient flexibility for future developments.

Finally, we are still some years out from the point where researchers with sophisticated search needs can expect the level of discoverability that they want/need?

Have I understood correctly?

KS: Yes, you are for the most part correct in summarizing my opinion.

A couple of small clarifications: We know from repository managers that the majority of users are coming to repositories from commercial search engines and not through harvesting services or the search facility built into the repository; and we know from [user studies](#) that the starting point to find information for many researchers is through Google or Google Scholar.

Currently, as things stand, the content in repositories is not highly ranked in Google Scholar, and in terms of Google, repositories are indexed alongside billions of other pages. So, no, this is not ideal for the discoverability of repository content, particularly for key word or topic-based searching.

I note that in the early days of Google Scholar, the open access community advocated for the search results to be tagged as open access (or not). Obviously we were not successful, but this would have enabled users to limit results to open access content and certainly been a boost for the visibility of repository content in this context.

I do believe the discoverability of repository content will improve greatly in the coming years. Refining the cross-repository search services, those that are based on harvested metadata, will depend on improving the standardization and comprehensiveness of metadata records. Technology will help with this. There are new, automated methods for assigning metadata and repository software platforms can build-in standard vocabularies and metadata elements.

The greater challenge is coming to an agreement about common terminologies and approaches across the entire repository community. COAR will play an important role by acting as a forum whereby the repository community can make these kind of collective decisions.

There will also likely be a number of services developed in the coming years to facilitate full-text searching through harvesting the content. According to [Petr Knoth](#) (Knowledge Media Institute, The Open University, UK) who has been doing research in this area through the CORE initiative referenced earlier, there still are a [number of technical and legal barriers](#) to full text harvesting from repositories.

However, in the coming years, I expect that the repository community will begin to address these barriers, especially the technical ones.

Again, I hope that COAR can play a role in developing solutions and disseminating best practices.

SHARE or CHORUS?

RP: You said (or at least implied) that repositories should be viewed as tools to enable the research community to “assume greater responsibility for managing, providing access and preserving the content created through research”. And you cited [SHARE](#) as an example of an initiative focussed on providing interoperability between repositories.

*It is worth noting that SHARE is a response by librarians to the **OSTP Memorandum**, which directs US Federal agencies to develop plans to ensure that the published results of research they have funded is made OA. As such, SHARE could be viewed as a good example of how research institutions can try to take greater responsibility for scholarly communication, since it would put librarians in charge of managing access to papers released as a result of the OSTP Memorandum.*

*However, you will know that publishers have proposed an alternative model based on **CHORUS**. The aim of CHORUS is to ensure that it is publishers rather than librarians who manage access to these papers, and it demonstrates their wish to remain firmly in control of scholarly communication, even after research papers have been made OA.*

How would you respond to someone who argued the following: Since the research community is finding it difficult to fill repositories (a point frequently made, not least by the Finch Report), and both difficult and time-consuming to create the necessary infrastructure to ensure repository content is optimally discoverable, might it not make more sense to outsource the task to publishers via initiatives like CHORUS? After all, CHORUS will deliver OA, and since publishers have greater resources they might be expected to undertake the task more effectively, and more quickly. Moreover, since it is they who publish the papers in the first place, they already have all the content in place.

KS: My major concern about CHORUS is that the publishing community would have too much control of the scholarly communication system. A number of large publishers have already demonstrated that they don't support the principle of open access (remember **PRISM**).

Frankly, the interests of publishers often lie elsewhere and they may be motivated by things such as profit margin not the public good.

On the other hand, at the core of the mission of the university and the library is the advancement and dissemination of knowledge. It seems to me that the world's collective knowledge created through research should rest in the hands of long-term actors whose *raison d'être* is to ensure that it is preserved and remains accessible to all.

CHORUS may seem like an appealing option for the US agencies at the moment, but the long-term implications are that the research community will have little control or ability to influence the future directions of scholarly communication if we take that route.

I'm also very concerned about the costs of such a system. Article processing fees are already way too high for many researchers, especially in developing countries. The **recent study** of APCs undertaken by the Wellcome Trust and others found that the average per article APC is \$1,418 USD for open access publishers. I don't believe this can scale globally and will ultimately result in disadvantaging a large number of researchers who can't afford to pay.

RP: *You are right that speed and effectiveness is one thing, cost and ownership something else. And as you suggested earlier, if the research community were to take greater responsibility for managing access to research it could hope to "alleviate some of the inflationary aspects of scholarly publishing and enable us to have more influence on future directions."*

This reminds me of what your colleague Eloy Rodrigues said to me last year. The future of scholarly communication, and its cost to the research community, he suggested, will depend on whether there is a "research-driven" transition to open access or a "publishing-driven" transition (in other words, whether the transition prioritises the needs of the research community or the needs of publishers). I would think that the competing SHARE and CHORUS initiatives are representative of these two approaches, and this suggests to me that in the coming years we will see publishers and librarians jostling for control of the scholarly communication system. And if that is right, the institutional repository will surely become a key battleground in the struggle.

Would you agree? And if it wants to ensure a "research-driven" transition to OA what should the wider research community be doing in your view?

KS: The choices that institutions make now about how they are going to invest in scholarly communications are absolutely critical.

First of all, I think the Green Road is key. We must collectively build and maintain a global system of repositories. It introduces competition into the system and will act as an important deterrent to arbitrary price increases by publishers.

It will also demonstrate the important role that institutions play in the stewardship of research outputs. To that end, institutions should devote more resources to their repository operations in order to improve repository services and increase the size of their collections.

Secondly, we should encourage and sponsor the development of new publishing models and value-added services that conform to our vision.

In terms of repositories, this would include better cross-repository discovery services, text mining capabilities, disciplinary views, and the development of overlay journals. Leslie Chan, for example, *makes the case* that the distinctions between “journal” and “repository” are increasingly blurred and that “mega-journals” are essentially repositories with overlay services.

We should be participating in projects that demonstrate the added value of repositories and repository networks across the research life cycle. Of course, this will require that we take some risks, which is a difficult case to make in hard economic times to (often) risk adverse organizations.

Global discussion

RP: You said that the way in which scholarly communication develops will vary “according to content type, discipline, and region.” Certainly, as OA develops we do appear to be seeing distinctive regional differences emerging. For instance, where the pay-to-publish gold OA model is being pushed heavily by the UK and The Netherlands there is still more of a focus on green OA in North America. Meanwhile, in Africa and Latin America a repository-based publishing model currently appears to dominate.

As things stand I would expect to see the Global North increasingly move to a pay-to-publish gold OA model and the Global South to a free-to-publish/free-to-read repository-based publishing model similar to that pioneered by SciELO and AJOL. If that proves the case, however, will it be the best outcome in a global research environment?

When I spoke to Dominique Babini last year she said “[W]e owe ourselves a global discussion about the future of scholarly communication”. And she added, “Now that OA is here to stay we really need to sit down and think carefully about what kind of international system we want to create for communicating research, and what kind of evaluation systems we need, and we need to establish how we are going to share the costs of building these systems.”

This would seem to imply a more global approach than we are currently seeing develop. Would you agree with Babini? If so, who should organise the global discussion she has called for, and who should take part in it?

KS: Yes, I agree, and I would add that we should consider carefully the unintended consequences of adopting the various models.

“What kind of system do we want to create for communicating researcher?” I would propose that we want one in which all researchers can access and contribute to, regardless of geographic location or discipline; and where the knowledge created is assessed on its real value, rather than on the region from which it emerges or the so called “impact” of the journal in which it is being published.

A dual system as you describe above is not ideal and I believe it will create inherent inequalities across the regions. Especially if we continue to rely on impact measures that do not reflect the quality of the research, but rather serve to prop up the traditional publishing system.

I believe there is a general lack of awareness in the “north” about the “southern” perspective and that we do need to ensure that the voices from the

south are heard.

In terms of the global discussion, we already have a number of international forums for exchange: the funding agencies have the [Global Research Council](#); libraries have organizations such as the [SPARCs](#) and [IFLA](#); the repository community has COAR; and, publishers have their own venues.

UNESCO, and the governments represented there, has also become interested in open access. We could begin the global discussion by facilitating greater dialogue across these different stakeholder organizations.

One missing but very important link is the research community. It's clear that many researchers have not been sufficiently engaged with the issues of open access to understand the nuances. For example many researchers still equate open access with open access journals. So we need a mechanism for bringing those communities into the discussion as well.

It is illuminating to note that a parallel global discussion is currently occurring in the area of research data through the Research Data Alliance ([RDA](#)). It has been comparatively easy in the context of research data to bring together the key stakeholders – researchers, data repositories, institutions, and funding agencies – to adopt a common vision and agree on practical strategies for moving forward.

Why haven't we been able to do that for publications? The essential difference is that for publications, there are some parties that have a significant financial interest in maintaining control of the system. This makes the global discussion far more challenging.

RP: Thank you very much for taking the time to speak with me.

Posted by Richard Poynder at [12:15](#)



5 comments:



Stevan Harnad said...

THERE'S NOTHING WRONG WITH GREEN OA REPOSITORIES THAT EFFECTIVE GREEN OA MANDATES WON'T CURE

Kathleen Shearer is right that the Green Road is the *key* -- but effective Green OA mandates are the *motor*.

Repositories are near empty. Repository functionality can always be improved, but no improvement of repository functionality will provide their missing content. That content will only be provided (by the researchers who produce the research) if the researchers' institutions and funders require (mandate) that they provide it, immediately upon acceptance for publication, as a prerequisite for research performance evaluation and funding.

There are currently well over 3000 repositories worldwide but fewer than 300 Green OA mandates worldwide, and many of them are weak, ineffective mandates (compare [ROAR](#) and [ROARMAP](#)).

What needs to be done on now is (1) for the institutions and funders that have already adopted Green OA mandates to upgrade to what has proved to be the strongest and most effective mandate model (Liège/HEFCE) and (2) for the many remaining institutions and funders adoption have not yet mandated Green OA self-archiving to likewise adopt the Liège/HEFCE model.

Until then, COAR's mission to "enhance the visibility and application of research outputs through a global network of open access digital repositories" will remain unfulfilled and unfulfillable.

See:

**The Liège ORBi model: Mandatory policy without rights retention
but linked to assessment processes.**

**The only way to make inflated journal subscriptions unsustainable:
Mandate Green Open Access**

May 05, 2014 11:35 am 🗑️

Stuart Shieber said...

Richard Poynder raises with Kathleen Shearer the issue of “dark” deposits in [Harvard’s DASH repository](#). He implies that the presence of a subset of articles in which the deposited article is not made available is a grave failing.

Ms. Shearer’s response is exactly right: “I’m not sure about the specific case of DASH, but this really speaks to the collection policy of the individual repository.” I’ve explained DASH’s collection policy with respect to dark deposits in some detail in my 2011 post “[The importance of dark deposit](#)”. In a nutshell, part of the role of the repository is an archival one – to collect the research output of the institution as broadly as possible. We therefore don’t turn articles away. But we also don’t distribute articles from DASH when we don’t hold rights to do so or when authors for whatever reason request us not to. (The particularly unrepresentative case of Professor Knoll’s large number of dark deposits is an instance of the latter. We do not, as a matter of principle and policy, unilaterally override the wishes of authors.)

I believe our collection policy – to deposit articles into DASH even if we cannot (yet) distribute them by right or author preference – is reasonable, and in fact preferable to policies that disallow dark deposits. I won’t rehearse [the seven reasons why](#), though I especially commend Reason 5 to the interested. The best evidence that we are doing something right is that the over 17,000 articles in DASH have been downloaded almost 3.2 million times, and at an increasing pace. Fixation on the subset that we avoid distributing in deference to legal or moral rights seems to miss the point.

May 09, 2014 4:14 pm 🗑️



Richard Poynder said...

@Stuart: I appreciate your taking the time to comment. I did not intend to imply that Harvard is guilty of a grave failing, and I do not believe I am fixated.

My objective in the Q&As I undertake is to draw out some of the many issues that surround OA. In the case of the comments that you refer to my aim was to air a topic concerning OA repositories that many puzzle over, and seems to me to be something deserving of discussion. As I say, thank you for responding.

Harvard describes DASH as a “[central, open-access repository of research by members of the Harvard community](#)”.

In that context, I made the following points:

1. The DASH repository is widely viewed as (and promoted by Harvard as) a poster child of the OA movement.
2. From what Kathleen Shearer said I inferred she believes OA repositories should always provide access to the full text (as well as the metadata) of papers they showcase.
3. In any case, I think most people expect the full text of papers deposited in an OA repository to be both present and freely available to all.
4. Certainly DASH has been criticised for not providing free access to the full text of all the papers it contains (and I linked to one such criticism).
5. While the criticism I pointed to dates from several years ago DASH does today still contain details of papers for which it does not provide access to the full text (and I linked to five examples that I found at random).
6. Some of these papers do not provide a link to the full text, others provide a link to the publisher’s site, where the reader is asked to pay up to \$35 to view them. I suggested that this cannot be described as OA.

I understand your point about dark deposits. I believe the standard practice for dealing with such deposits is to provide a [Request Copy Button](#) in the repository so that researchers can automatically request that the author send them a copy. As I indicated, I could not find a Request Copy Button in DASH. Perhaps I missed it?

Congratulations on the number of downloads.

May 10, 2014 9:11 am 

Stuart Shieber said...

I apologize for my overstrong language ("grave", "fixating"). It's so hard to get tone right in comment threads.

We do refer to DASH as a "central, open-access repository of research by members of the Harvard community", and I think it is just that. Peter Suber's take on our use of the phrase "open-access repository" is trenchant I think:

"We call something a 'bookstore' even if it also sells magazines and greeting cards. We call something a 'grocery store' even if it also sells spatulas and pot holders. We call something a 'drama' even if it includes some comedy, and vice versa.

"An 'OA repository' may have some dark content without contradiction. The 'OA' in the name designates the primary purpose of the repository, not the exclusive purpose, just as with 'book' in 'bookstore' and so on.

"If a fuller description of a bookstore were 'store for books, magazines, greeting cards, mugs, and pens', then a fuller description for DASH would be 'repository for open access and preservation'. It's fair and commonplace to abbreviate these long descriptions into short names that leave out much of the descriptive nuance. If it's fair to say 'bookstore', then it's fair to say 'OA repository'."

By the way, the proportion of dark material in DASH is relatively small, about 10%, and we're looking into what portion of that might be "brightened".

May 12, 2014 7:48 pm 



Richard Poynder said...

Thank you for this further response Stuart.

You do not say why the DASH repository does not have a Request Copy Button.

May 19, 2014 12:48 pm 

[Post a Comment](#)

[Newer Post](#)

[Home](#)

[Older Post](#)

Subscribe to: [Post Comments \(Atom\)](#)

Website maintained by [NARKAN](#). Powered by [Blogger](#).